DEVELOPMENT OF SIGNAL SEGMENTATION TECHNIQUE AND IMPROVED FUZZY K NEAREST CENTROID NEIGHBOR (IFKNCN) CLASSIFIER FOR AUDIO IDENTIFICATION SYSTEM

HARYATI BINTI JAAFAR

UNIVERSITI SAINS MALAYSIA

2015

DEVELOPMENT OF SIGNAL SEGMENTATION TECHNIQUE AND IMPROVED FUZZY K NEAREST CENTROID NEIGHBOR (IFKNCN) CLASSIFIER FOR AUDIO IDENTIFICATION SYSTEM

by

HARYATI BINTI JAAFAR

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

September 2015

ACKNOWLEDGEMENT

First and foremost, all praises to Almighty ALLAH S.W.T. for delivering me the patience, strength and guidance in completing my PhD at Universiti Sains Malaysia. I would like to express my gratitude to my supervisor Dr. Dzati Athiar binti Ramli for her encouragement and great support. Her constant guidance, subjective criticism, continuous support and great dedication time are most valued to gain my confidence in achieving the objective of this project.

I would like to extend my gratitude to Dr. Bakhtiar Affendi bin Rosdi for his encouragement and thoughtful discussions. Special thanks to Dr. Shahriza bin Shahrudin of School of Pharmaceutical Sciences who helped me a lot during data collection process. My sincere thanks go to Institute of Postgraduate Studies (IPS), USM for the contribution made to support the research through the Postgraduate Research Grant Scheme (PGRS) No. 8046019, Research University Grant No. 814161 and Ministry of Education for granted me the myBrain fellowship as my financial support to complete my study.

My warmest feeling is addressed to my beloved parents, En. Jaafar bin Md. Yusof and Pn. Asmah binti Hedzir for their support. To my sisters, Hartini and Hidayani and my brother, Muhammad Ramlan, thank you for always give me encouragement to complete my PhD. I also want to express my special thanks to all the academic, administrative and technical staff of School of Electrical and Engineering for their kind support and assistance. My appreciation goes to my friends, Salwani Ibrahim, Najah Ghazali, Mastika Suhaila and post graduate roommates for their support, help and understanding throughout the period. Thanks a million.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CA	Classification Accuracy
CNN	Condensed Nearest Neighbor
COA	Centre of Area
COG	Centre of Gravity
dB	Desibel
DCE	Delta Cepstral Energy
DDCE	Delta-Delta Cepstral Energy
DFT	Discrete Fourier Transform
DRMS	Dimension Reduction of The Modulation Spectrum
E	Energy
ED	Euclidean Distance
EER	Equal Error Rate
FA	Factor Analysis
FCNN	Fast Condensed Nearest Neighbor
FEC	Front End Clipping
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FIS	Fuzzy Inference System
FkNCN	Fuzzy-Based k Nearest Centroid Neighbor
FkNN	Fuzzy k Nearest Neighbor
FL	Feature Line
FWHM	Full Width at Half Maximum
GA	Genetic Algorithm
GG	Gabriel Graph
GMM	Gaussian Mixture Model
Hz	Hertz
IBG	Intelligent Biometric Group
IDFT	Inverse Discrete Fourier Transform
IFkNCN	Improved Fuzzy-Based k Nearest Centroid Neighbor
kNCN	k Nearest Centroid Neighbor
kNN	k Nearest Neighbor

LPC	Linear Predictive Coding
LPCC	Linear Prediction Cepstral Coefficients
MAR	Multivariate Auto-Regression
MF	Membership Functions
MFCC	Mel Frequency Cepstral Coefficients
MLP	Multilayer Perceptron
MSC	Mid Speech Clipping
NBNN	Naïve-Bayes Nearest Neighbor
NCFL	Nearest Centroid Feature Line
NCN	Nearest Centroid Neighborhood
NDS	Noise Detected As Speech
NFL	Nearest Feature Line
NIR-LED	Near Infrared Light Emitted Diode
NN	Nearest Neighbor
PCA	Principle Component Analysis
PSO	Particle Swarm Optimization
RNG	Relative Neighborhood Graph
RNN	Reduced Nearest Neighbor
ROI	Region of Interest
SM	Sinusoidal Modelling
SN	Surrounding Neighborhood
SNR	Signal to Noise Ratio
STAZCR	Short Time Average Zero Crossing Rate
STE	Short Time Energy
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
SWF	Similarity-Weighted Function
USM	Universiti Sains Malaysia
WkNN	Weighted k Nearest Neighbor
WSF	Weighted Similarity Function
ZCR	Zero Crossing Rate

CHAPTER 1

INTRODUCTION

1.1 Overview of Audio Identification System

Audio signal plays a critical role in daily communication, perception of environment and entertainment. Today, the audio signal technologies have experienced a vigorous growth in a wide range of applications such as in the identification of animal sound, human speech and speaker and music genre (Stranneby and Walker, 2004, Rao, 2007). While the human auditory system is able to process the complex sound mixture, the human computer interaction can be extremely difficult and challenging. In spite of almost four decades of research, the design of the audio identification still remains as an elusive goal for the researchers to investigate.

In general, an audio identification system consists of five major phases as shown in Figure 1.1. The first phase concerns with the representation of input data from the objects to be recognized (Tohka, 2013). In the second phase, there are some tasks that can be included to improve the data quality such as the noise reduction, filtering, encoding and enhancement. Subsequently, the third phase is applied to separate the signal so that each of them can be represented as a classified object. These segmented signals are later mapped onto points in space during the feature extraction phase. Here, the dimensionality of the data is reduced by measuring and retaining certain characteristics of features (Burgers et al., 2009). The final phase involves the classification process where the data are assigned to their pattern classes.

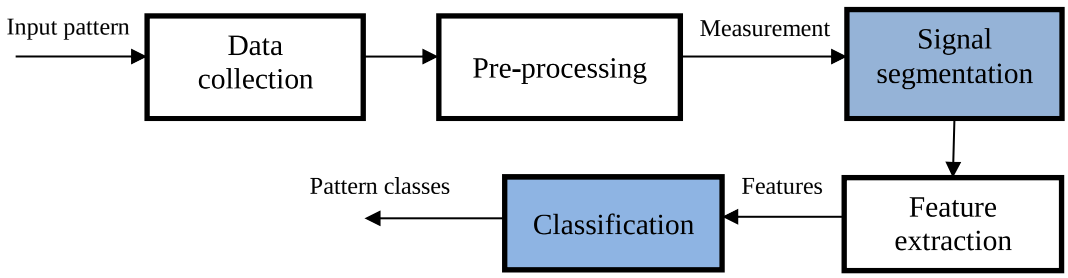

Figure 1.1 Audio identification system

This research focuses on two main aspects of the audio identification system which are signal segmentation and classification processes. They are shown in blue-coloured box in Figure 1.1. The aim of this research is to devise a signal segmentation technique and an appropriate classifier for the audio identification system. In order to evaluate the applicability and accuracy of the proposed methods, numerous experimental studies using audio databases have been conducted and the results have been compared, analyzed and discussed.

1.2 Problem Statement and Motivation

1.2.1 Signal Segmentation Process

Segmentation of the continuous signal is a fundamental task in any audio identification system. This process maintain the desired signal and remove the undesired signal, thus saving the computation cost and increases the accuracy of the whole system (Costa et al., 2012).

So far, there are many techniques of the signal segmentation have been proposed. The most commonly mentioned techniques are the Energy (E) and Zero Crossing Rate (ZCR) (Chen et al., 2012) and Sinusoidal Modelling (SM) (Harma,

2003) techniques. These techniques may be sufficient to segment the desired signal in a clean or very high Signal to Noise Ratio (SNR) conditions.

However, when the signal is corrupted by noise, it can be very hard to distinguish between the desired and undesired signals. Some of the noises are produced in the electronic devices such as the white and pink noises (stationary noise) which are easy to handle and can be filtered out in the pre-processing phase (Becchetti and Ricotti, 1999). It can be very challenging to segment the desired signal in more complex environments such as the background music, background speech, and reverberation which are commonly originated from the non-stationary noise. This is because the amplitude values of the undesired signal can be higher than the desired signal (Galleani and Cohen, 2006; Fujimoto, 2011).

Because real-life situation is generally dominated by many types background noise with low SNR, performance of signal segmentation degrades significantly, discouraging its practical use. Therefore, to overcome the challenges of signal segmentation becomes as the motivation in this research. In this research, problem encountered in stationary and non-stationary noises with low SNR (20dB, 10dB and 5dB) are addressed and solved.

1.2.2 Classification Process

Apart from signal segmentation, choosing the right classification algorithm is also important in an audio identification system. In the literature, two types of classifications namely the unsupervised and supervised classifications are commonly used (Sheskin, 2011). The unsupervised classification does not require any prior knowledge about the training set. In contrast, the classification of the samples of labelled classes (training set) during the classification stage is referred to as a supervised classification (Wu et al., 2008). Supervised classification is divided into parametric and non-parametric classifiers and the difference of these classifiers will be discussed further in Chapter 2.

The focus of this thesis is on the nonparametric classifier as it did not assume the sample distribution thus alleviating the complexity of building the models. Among the nonparametric classifier, k Nearest Neighbor (kNN) has been widely used for pattern classification. This classifier is simple to be implemented and it eases the classification process (Cover and Hart, 1967). This classifier is particularly well suited for multi-modal classes as well as applications in which an object can have many class labels which makes this classifier outperformed other non-parametric classifiers (Wu et al., 2008).

Nevertheless, there are some problems encountered that leads to the poor performance of the kNN especially in the large datasets. Firstly, the lacked of information on the samples distribution (Chaudhuri, 1996). Secondly is concerning the weighting issues in assigning the class label before classification (Wang et al., 2007; Imandoust and Bolandraftar, 2013). Finally is the computational time problem since all distances between a query object to the other training objects have to be taken into account in order to find the nearest neighbors for the query object (Wu, 2008).

Realizing the disadvantage encountered in kNN, this research is motivated to develop a new classifier called based on kNN that capable of solving the aforementioned problems. It is hope that the findings of this study will provide more insights in improving the classification algorithm in the different types of audio signal database.

1.3 Objectives of Studies

Based on the aforementioned problems in Section 1.2, the main objectives of this study are to develop an intelligent signal segmentation and classifier for audio identification system. They are achieved by considering the following sub-objectives:

i. To develop a signal segmentation process in the different background noises and low SNR by proposing a new technique based on the combination of the Short Time Energy (STE) and Short Time Average Zero Crossing Rate (STAZCR).
ii. To develop a classifier system based on Nearest Centroid Neighbor (NCN) classifier by proposing the Fuzzy-Based k Nearest Centroid Neighbor (FkNCN) classifier in order to solve the sample distribution and weighting issues. Subsequently, this classifier is improved by introducing the Improved Fuzzy-Based k Nearest Centroid Neighbor (IFkNCN) classifier in order to reduce the computational complexity in FkNCN.
iii. To evaluate the proposed STE and STAZCR techniques and IFkNCN into the different audio signal databases.

1.4 Scope of Research

The experimental studies for this research has been conducted on the benchmark and collected databases. The benchmark database is the audio-visual digit database (Sanderson and Paliwal, 2003). It is obtained by taking the information of a speaker's tone from a recorded speech and inflection analysis. The collected database is a comprehensive collection of frog calls from 15 frog species recorded from the Malaysian forest.

In the audio signal segmentation process, two techniques namely the STE and STAZCR which based on short-time analysis are proposed. Both STE and STAZCR are combined together to determine the start and end point detections to detect the desired signal. At the same time, it should be able to exclude the background noise and undesired signal. The noises included stationary and non-stationary noises which are sourced from the man-made and natural environments.

During the classification process, in the FkNCN, a surrounding-fuzzy based rule is firstly investigated. This rule incorporates centroid-based distance and fuzzy rule to solve the sample distribution and the ambiguity of the weighting distance between the query point and training samples.

Subsequently the development on the IFkNCN is implemented. The development of this classifier is divided into two stages which are the building and searching stages. During the building stage, fuzzy inference system is employed to set a threshold. The training sample that is located far from the query point or outside the threshold is called as a noisy sample, thus not fitting in the assumed class label for the query point. By removing the noisy sample, the future processing can mainly focus on the important training samples and therefore the computational complexity can be decreased. Similar procedure in the searching stage of the FkNCN is used for the development of the IFkNCN whereby the surrounding fuzzy-based rule is implemented before doing the classification of the query point. The overall architecture of this thesis is illustrated in Figure 1.2 where the green-coloured box shows the combination of this thesis.

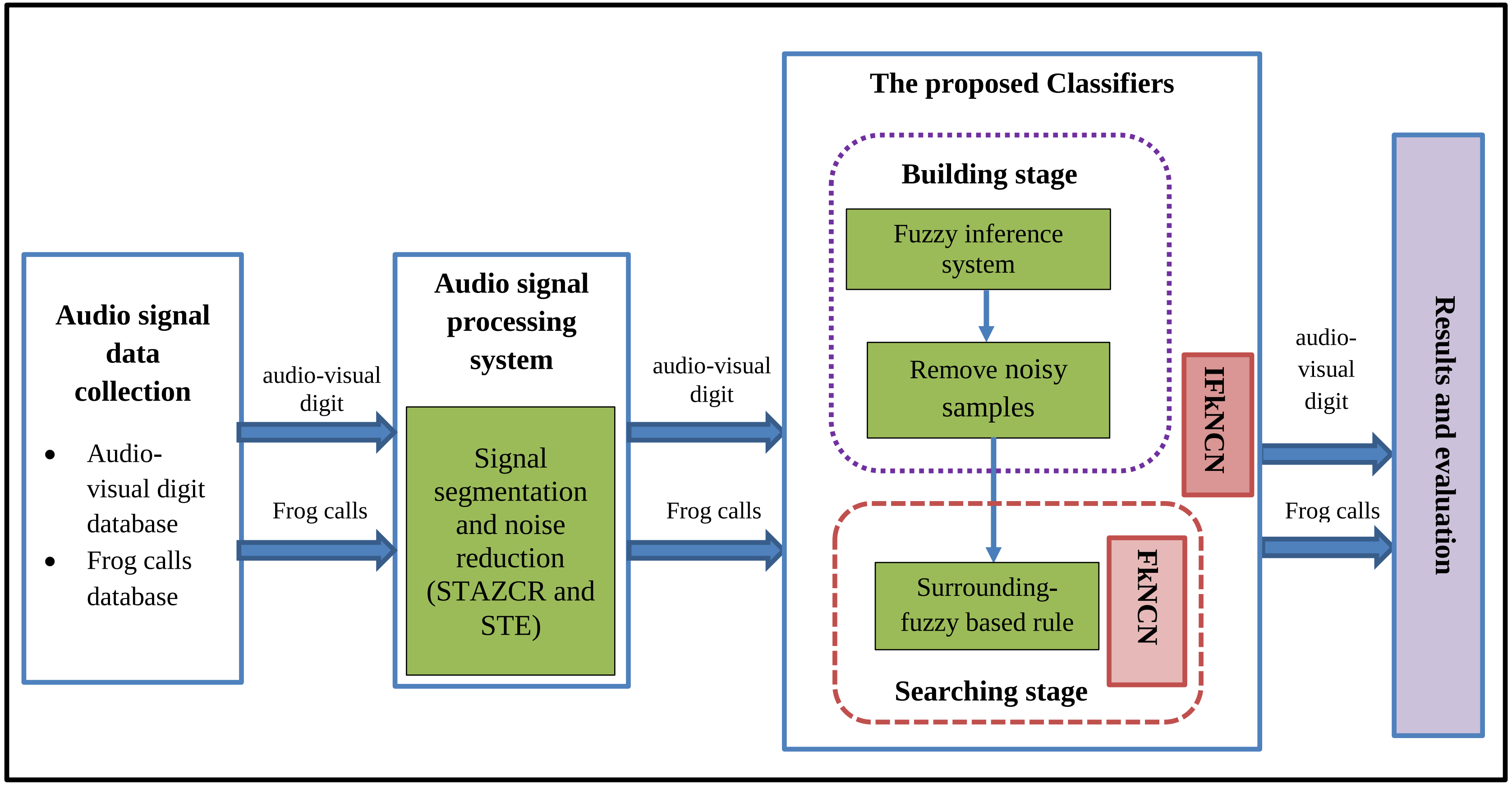

Figure 1.2 Overall architecture of research study

1.5 Thesis Contribution

This research has contributed a few discoveries as stated below:

i) An improvement of the signal segmentation technique is introduced by implementing the Short Time Energy (STE) and Short Time Average Zero Crossing Rate (STAZCR) techniques in this research.

ii) A new classifier called as Fuzzy-Based k Nearest Centroid Neighbor (FkNCN) classifier is first proposed in this research. In this classifier, the query point is classified based on the concept of the surrounding-fuzzy based rule in which the selected training sample is defined as the Nearest Centroid Neighbor (NCN). The NCN is selected based on the information of the training samples distribution around the query point. The query point is classified by solving the ambiguity of the weighting distance between the query point and its NCNs.

iii) An extension of the FkNCN classifier which is the Improved Fuzzy-Based k Nearest Centroid Neighbor (IFkNCN) classifier is proposed to reduce the computational complexity by removing the noisy sample that that is located outside the threshold.

1.6 Thesis Outlines

This thesis is organised in six chapters. The overviews of audio identification systems, signal segmentation and classification process are covered in Chapter 1. The problem statement, objectives, scope of research and thesis contribution are also presented here.

Each stage that is involved in the development of an audio identification system such as the audio signal processing systems, signal segmentation and classification process are reviewed as written in Chapter 2. Previous works on the signal segmentation and classification process are also reported in this chapter.

The methodology of the developed and proposed technique in signal segmentation is stated in Chapter 3. The performance evaluation of the signal segmentation is stated in this chapter.

The algorithms of the proposed FkNCN and IFkNCN classifiers are presented in Chapter 4. The difference between the FkNCN and IFkNCN classifiers will be analysed in this chapter.

The results obtained by using the techniques and classifiers that are mentioned in Chapter 3 and 4 respectively are covered in Chapter 5. It consists of three parts in which the first part describes the details of databases that are used in this research. The second part is the result of the proposed signal segmentation techniques based on subjective and objective evaluations and the classification accuracy. The results of the proposed technique are later compared to the other signal segmentation techniques. The second part is the classification results using the proposed classifiers. There are several states of art in the other classifiers are analyzed and compared to the proposed classifier.

Finally, a conclusion for this study is made and the notable contributions to this work are highlighted in Chapter 6.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter devotes on the fundamental theory and literature studies of audio processing and classifier. These fundamentals will be used in the proposed methodology in Chapter 3 and Chapter 4. This chapter is organized in six sections. Section 2.2 presents the information on the audio processing that includes the pre-processing, signal segmentation and feature extraction phase. Subsequently, the review of the various signal segmentation techniques is discussed in Section 2.3. Section 2.4 reviews the classification process involving the supervised an unsupervised classification that has been used in the previous literature. The review of the Nearest Neighbor (NN) classifier and the improvement of this classifier is presented in Section 2.5. Section 2.6 presents the review of a Support Vector Machine (SVM) classifier. Finally, Section 2.7 presents the summary of this chapter.

2.2 Audio Signal Processing

The pre-processing, signal segmentation and the feature extraction phases are important to improve the effectiveness of an audio signal processing (Theodoridis and Koutroumbas, 2006). Each of the collected data undergoes a series of audio signal processing which are proceeded in six processes i.e. digitization, pre-emphasis, framing, windowing, signal segmentation and feature extraction as shown in Figure 2.1. Each process is discussed in details in the following subsections.